Autoregressive Models for Capture-Recapture Data: A Bayesian Approach

Devin S. Johnson* **and Jennifer A. Hoeting**
Department of Statistics, Colorado State University
Fort Collins, CO 80523
**email:* johnson@stat.colostate.edu

May 3, 2002

Summary

In this paper, we incorporate an autoregressive time-series framework into models for animal survival using capture-recapture data. Researchers modelling animal survival probabilities as the realization of a random process have typically considered survival to be independent from one time period to the next. This may not be realistic for some populations. Using a Gibbs sampling approach we can estimate covariate coefficients and autoregressive parameters for survival models. The procedure is illustrated with a waterfowl band recovery dataset on Northern Pintails (*Anas acuta*). The analysis shows that the second lag autoregressive coefficient is significantly less than 0, suggesting that there is a triennial relationship between survival probabilities and emphasizing that modelling survival rates as independent random variables may be unrealistic in some cases. Software to implement the methodology is available at no charge on the internet.

Key words: Autoregressive models; Bayesian inference; MCMC; Survival estimation; winBUGS.

1 Introduction

The investigation of factors that affect animal survival has become an increasingly important aspect of ecological research (Lebreton et al., 1992). It is often of interest to account for survival rates with covariates such as age, time, or weather factors (Buckland et al., 2000). Researchers have recently begun to explore the view that survival probabilities are realizations of a random process rather than fixed constants (Barry et al., 2001; Burnham, 2000; Burnham and White, 2002). Modeling survival probabilities via random effects allows one to account for over-dispersion (Barry et al., 2001) and unobserved (or random) environmental

factors (Burnham, 2000). To this point, however, realizations of the survival process have been considered to be independent from one time period to the next. In some situations, this may be an unrealistic assumption. For example, survival at weekly intervals over the course of one season would likely be correlated, or high survival in one period may lead to low survival in following periods due to lack of resources. Therefore, it is reasonable to consider a time series correlation structure, such as an autoregressive structure (AR), in models where survival is considered a random process.

Vounatsou and Smith (1995), Brooks et al. (2000a; 2000b), and Poole and Zeh (2002) have used Bayesian methods to estimate individual survival rates. Recently, Brooks et al. (2002) and Barry et al. (2001) have used Bayesian methods to estimate survival models with independent random effects as a way to model overdispersion. Burnham (2000) and Burnham and White (2002) have considered random effects in a non-Bayesian framework. Estimation via maximum likelihood is difficult in the context of random effects models. The likelihood is constructed by integrating over the random effects, and thus, an integration must be performed over all of the random effects included in the model for each iteration of an optimization algorithm. This same difficulty is encountered in generalized linear models (Zeger and Karim, 1991).

The use of random effects allows for modeling survival in a capture-recapture model via an AR process. The Bayesian paradigm provides several advantages over maximum likelihood estimation. Using Markov Chain Monte Carlo (MCMC) procedures (Robert and Casella, 1999), point estimates can be produced by sampling from the posterior distribution of the parameters. In addition, Bayesian methods allow for estimation of the unobserved random effects as well. For example, survival probabilities can be estimated for each individual time period. This is not feasible with maximum likelihood estimation procedures when random effects are included.

We consider models for two common types of capture-recapture data: open population mark recapture (MR), where animals are recaptured and released alive, and band return (BR), where animals are recovered dead after each hunting season. For each of these data types we develop the theoretical construction and estimation procedures for a m^{th} order autoregressive, AR(m), random effects model using a Bayesian approach. The Bayesian

analysis is illustrated using a long term waterfowl band recovery data set for Northern Pintails (*Anas acuta*).

2 Likelihood for Capture-Recapture Data

The likelihoods for open population mark-recapture (MR) data and band recovery (BR) data are structurally identical, the only major difference being a slight modification of the parameters. A complete description of the likelihood for capture-recapture data is given in Lebreton et al. (1992) and Brownie et al. (1985). We give a brief description here for the Bayesian methods presented in the next section.

Data are typically observed as an upper triangular array, $\mathbf{m}$, where the i, j^{th} element, m_{ij}, is the number of animals released at time t_i and subsequently recaptured (or reported, in the case of BR models) at time t_j (see Table 1, for example). The value I represents the number of capture occasions in which marking or banding is performed and J is the number of occasions in which recording recaptures or recoveries occurs. In MR studies, typically, $J = I$, while for BR studies, J may be greater than I due to the fact that marked animals may be harvested and reported after marking has stopped. Another component of the data is the $I \times 1$ vector $\mathbf{R} = [R_i]$, which contains the number of marked, or banded, animals released at each capture occasion. Each row of $\mathbf{m}$ is then modelled as a multinomial random variable with R_i trials and cell probabilities determined by survival probabilities and recapture or recovery probabilities. When using capture histories summarized into the sufficient statistics $\mathbf{m}$ and $\mathbf{R}$, the assumption that individuals have identical survival probabilities and recapture/recovery rates is assumed.

2.1 Open population mark-recapture likelihood

The MR model for survival estimation is also referred to as the Cormack-Jolly-Seber model (Cormack, 1964). This model is designed for studies in which a captured animal is labelled with a unique marking and released back into the wild population. At some point in the future, the marked animal may be recaptured, recorded, and released once again into the population. In MR studies, the first possible recapture of an animal marked at t_i is t_{i+1}, therefore, for these models $i = 1, \ldots, I$ and $j = i+1, \ldots, J+1$. Under the assumptions that individuals are independent and capture does not affect survival or recapture probabilities,

the resulting product multinomial likelihood is

$$\mathcal{L}(\boldsymbol{\phi}, \mathbf{p}; \mathbf{R}, \mathbf{m}) = \prod_{i=1}^{I} \binom{R_i}{m_{i,i+1}, \ldots, m_{i,J+1}, v_i} \xi_i^{v_i} \prod_{j=i+1}^{J+1} \left\{ \phi_i p_j \prod_{k=i+1}^{j-1} \phi_k (1 - p_k) \right\}^{m_{ij}}, \quad (1)$$

where, ϕ_i is the probability that an animal survives from capture occasion t_i to t_{i+1}, $i = 1, \ldots, I$ given that it is alive at t_i, and p_j, is the probability that an animal alive at t_j, $j = 2, \ldots, J+1$, is captured at t_j. The probability that an animal is never recaptured after release at t_i is given by

$$\xi_i = 1 - \sum_{j=i+1}^{J+1} \phi_i p_j \prod_{k=i+1}^{j-1} \phi_k (1 - p_k)$$

and $v_i = R_i - \sum_{j=i+1}^{J+1} m_{ij}$ is the number of animals captured at t_i and never subsequently recaptured during the study. In this section and for the remainder of this paper, a reverse order product is set equal to 1. For example, if $j = i + 1$, then $\prod_{k=i+1}^{j-1} \phi_k (1 - p_k) = 1$.

2.2 Band recovery likelihood

Band recovery models are designed for studies in which animals are captured, marked and released. Animals are then reported to the banding agency after harvesting by hunters. Therefore, at "recapture" occasions, the marked animals are removed from the population. The structure of the data remains in the **R**, **m** format, so, the form of the likelihood is the same as (1), the only modifications being a change in the cell probabilities of the multinomial distribution and ranges for the i, j indices. Since an animal can be harvested and reported in the same time period in which it was banded, the index ranges are set at $i = 1, \ldots, I$ and $j = i, \ldots, J$. The resulting likelihood is

$$\mathcal{L}(\boldsymbol{\phi}, \boldsymbol{\lambda}; \mathbf{R}, \mathbf{m}) = \prod_{i=1}^{I} \binom{R_i}{m_{ii}, \ldots, m_{iJ}} \xi_i^{v_i} \prod_{j=i}^{J} \left\{ \lambda_j \prod_{k=i}^{j-1} \phi_k \right\}^{m_{ij}}, \quad (2)$$

where λ_j is the probability that a marked animal, alive at t_j, is harvested between time t_j and t_{j+1} and reported to the banding agency. In the band recovery model $\xi_i = 1 - \sum_{j=i}^{J} \lambda_j \prod_{k=i}^{j-1} \phi_k$ and $v_i = R_i - \sum_{j=i}^{J} m_{ij}$. Notice, in the BR model, that since an animal can be reported in the same time period as marking, the i, i cell probabilities involve only the λ_i parameter. Therefore, there are only $J - 1$ survival probabilities even though there are J years of data.

3 A Bayesian Approach for AR(m) Survival Models

3.1 Model specification

We consider a generalized linear model for the probability that an animal survives from time t_j to time t_{j+1} of the form

$$g(\phi_j) = \mathbf{X}_j'\boldsymbol{\beta} + \epsilon_j, \qquad j = 1, \ldots, J, \tag{3}$$

where g is an appropriate link function to constrain survival between 0 and 1, $\mathbf{X}_j$ is a $P \times 1$ matrix of covariates collected at capture occasion j, $\boldsymbol{\beta}$ is a $P \times 1$ vector of regression coefficients, and $\boldsymbol{\epsilon} = (\epsilon_1, \ldots, \epsilon_J)' \sim N(\mathbf{0}, \boldsymbol{\Sigma})$. The covariance matrix, $\boldsymbol{\Sigma}$, can be any general form. Here we consider an AR(m) model which implies that the ϵ_j error terms are realizations from the stochastic process

$$\epsilon_j = \sum_{k=1}^{m} \rho_k \epsilon_{j-k} + z_j, \qquad j = 1, \ldots, J, \tag{4}$$

where $z_j \sim$ i.i.d. $N(0, \sigma^2)$ and $\boldsymbol{\rho} = (\rho_1, \ldots, \rho_m)$ is a set of parameters.

We assume the process represented by (4) is stationary. For this model, stationarity implies that the covariance between two survival probabilities is a decreasing function of the distance between two time points and is independent of any one time point. The stationarity assumption of the error process imposes a constraint on $\boldsymbol{\rho}$ such that the roots of the characteristic equation,

$$x^m - \rho_1 x^{m-1} - \cdots - \rho_m = 0,$$

must be less than 1 in absolute value (Harvey, 1993, pg 20). In terms of the parameters, an AR(1) process is stationary if $|\rho| < 1$, while, an AR(2) process is stationary if $|\rho_1| < 2$ and $1 < \rho_2 < 1 - |\rho_1|$.

By including random error terms in (3), we account for unknown environmental influences that might affect the probability of survival. Without the addition of the error terms, it is assumed that the covariates completely determine survival. Allowing for correlation between the random error terms in (4) provides the added complexity that unknown environmental conditions may be similar for capture periods close together in time, so, survival probabilities should also be related. Negative values for some elements of $\boldsymbol{\rho}$ might imply some density

dependent effects in the population. A year in which survival is above average may lead to a below average survival rates in the subsequent years due to lack of resources.

A stationary AR(m) model provides either positive or negative correlation between survival probabilities that decreases with an increasing separation in time. So, the AR(m) model provides the type of relationship between survival probabilities that is desired. In addition, the model is relatively straightforward. Lindsey (1999, pg 106) notes that for short repeated measurement studies, elaborate time-series modeling is not necessary or possible and a simple AR process is usually adequate. The vast majority of capture-recapture datasets are no more than 50 years long (Franklin et al., 2002). Therefore, capture-recapture data certainly fit into the category of short time series data.

The model specified in (3) is one where the time series component appears in the error term. In other AR model formulations, the time series component appears with the mean term (Lindsey, 1999). However, we prefer model (3) for ease of biological interpretation. Often, the goal is to determine what covariates best model survival probability. If all of the variation in survival probability is not accounted for with the covariates sampled, only then would it be advisable to determine what associations exist between survival probabilities and different time periods.

3.2 Bayesian parameter estimation

We adopt a Bayesian approach for estimating the parameters for an AR(m) capture-recapture models specified in Section 3.1. The goal of this approach is to estimate the posterior distribution of the parameters to make inference about the parameters and ecological hypotheses. This approach is relatively simple in comparison to maximum likelihood estimation (MLE). To estimate the parameters via MLE, it is necessary to evaluate the integrated likelihood of the form

$$\mathcal{L}(\boldsymbol{\phi}, \cdot, \sigma^2, \boldsymbol{\rho}; \mathbf{R}, \mathbf{m}) = \int_{\boldsymbol{\epsilon}} \mathcal{L}(\boldsymbol{\phi}, \cdot; \mathbf{R}, \mathbf{m}) N(\boldsymbol{\epsilon}; \mathbf{0}, \boldsymbol{\Sigma}) d\boldsymbol{\epsilon}$$

where, $\mathcal{L}(\cdot)$ is given by (1) or (2) and $N(\cdot)$ follows from (3) and (4). Therefore, for each step in an optimization algorithm a high dimensional integration must be performed. An alternative approach, quasi-likelihood (McCullagh, 1983), has been developed for random effects models in the generalized linear model setting. This approach involves the development of estimating

equations that behave like likelihood functions and hence often have the same properties. In the capture-recapture setting, however, the fact that the cell probabilities are functions of the survival probabilities makes quasi-likelihood estimation difficult as well. In the Bayesian paradigm, the unobserved random effects are treated as random variables along with the parameters and the integration is performed stochastically through a Markov chain which samples from the joint conditional distribution of the parameters and the random effects given the data. From this joint conditional distribution we can obtain point estimates and confidence intervals for the parameters of the model.

In what follow, we present a general estimation procedure for both mark recapture and band recovery data. To simplify notation, we will use the notation $\mathbf{r}$ to represent either the vector of capture probabilities, $\mathbf{p}$, or band return rates, $\boldsymbol{\lambda}$, depending on the type of data being considered. We will also use the j index range $1, \ldots, J$ for both MR and BR data as this will not change the estimation procedure. The observed data, $\mathbf{m}$ and $\mathbf{R}$, as well as the covariates, $\mathbf{X}$, will collectively be denoted by D.

We assume that the parameters $\boldsymbol{\beta}$, σ^2, $\boldsymbol{\rho}$, and $\mathbf{r}$ are independent *a priori*. The posterior distribution of the parameters and random effects is then given by

$$\begin{aligned} \pi(\boldsymbol{\beta}, \sigma^2, \boldsymbol{\rho}, \boldsymbol{\epsilon}, \mathbf{r}|D) \quad \propto \quad & \mathcal{L}(\boldsymbol{\beta}, \boldsymbol{\epsilon}, \mathbf{r}; D) \times \ |\boldsymbol{\Sigma}|^{-1/2} \exp\left\{-\boldsymbol{\epsilon}'\boldsymbol{\Sigma}^{-1}\boldsymbol{\epsilon}/2\right\} \\ & \times \pi(\boldsymbol{\beta})\pi(\sigma^2)\pi(\boldsymbol{\rho})\pi(\mathbf{r}). \end{aligned} \tag{5}$$

In order to draw a sample from this distribution we will make use of the Gibbs sampler (e.g. Section 2.1 of Chen et al. 2000), which requires the full conditional distributions for each of the parameters. A sample from the joint posterior distribution can be drawn by successively drawing from the full conditional posterior distributions for each of the parameters.

A simplification of the full conditional distributions results from the fact that the likelihood function can be broken into three parts. In addition, since the posterior is only defined up to a multiplicative constant, we can ignore the multinomial coefficients. Therefore, the likelihood can be rewritten as

$$\mathcal{L}(\boldsymbol{\beta}, \boldsymbol{\epsilon}, \mathbf{r}; D) \propto V \times \mathcal{L}_{\phi} \times \mathcal{L}_{\mathbf{r}},$$

where for both MR and BR data, $V = \prod_{i=1}^{I} \xi_i^{v_i}$. For MR data

$$\mathcal{L}_\phi = \prod_{i=1}^{I} \prod_{j=i+1}^{J+1} \left(\phi_i \prod_{k=i+1}^{j-1} \phi_k \right)^{m_{ij}}$$

and

$$\mathcal{L}_{\mathbf{r}} = \prod_{i=1}^{I} \prod_{j=i+1}^{J+1} \left\{ p_j \prod_{k=i+1}^{j-1} (1 - p_k) \right\}^{m_{ij}},$$

and for BR data

$$\mathcal{L}_\phi = \prod_{i=1}^{I} \prod_{j=i}^{J} \left\{ \prod_{k=i+1}^{j-1} \phi_k \right\}^{m_{ij}}$$

and

$$\mathcal{L}_{\mathbf{r}} = \prod_{i=1}^{I} \prod_{j=i+1}^{J} \lambda_j^{m_{ij}}.$$

Now, with the partitioned form of the likelihood we can simplify the full conditional distributions for each of the parameters. Due to the fact that all but one parameter has a nonstandard distribution, we will only give the conditional distributions up to a proportional constant.

If the regression parameters for the covariates, $\boldsymbol{\beta}$, in (3) are independent *a priori*, then the full conditional of the coefficient for the l^{th} covariate, β_l is given by

$$f(\beta_l | \boldsymbol{\beta}_{-l}, \sigma^2, \boldsymbol{\rho}, \boldsymbol{\epsilon}, \mathbf{r}, D) = f(\beta_l | \boldsymbol{\beta}_{-l}, \boldsymbol{\epsilon}, \mathbf{r}, D) \propto V \cdot \mathcal{L}_\phi \cdot \pi(\beta_l) \qquad l = 1, \ldots, P.$$

Likewise, independent priors for the components of $\mathbf{r}$, the vector of capture probabilities for MR data or band return rates for BR data, give

$$f(r_l | \mathbf{r}_{-l}, \boldsymbol{\beta}, \boldsymbol{\epsilon}, \sigma^2, \boldsymbol{\rho}, D) = f(r_l | \mathbf{r}_{-l}, \boldsymbol{\beta}, \boldsymbol{\epsilon}, D) \propto V \cdot \mathcal{L}_{\mathbf{r}} \cdot \pi(r_l) \qquad l = 1, \ldots, J.$$

When deriving the full conditional distribution of ϵ_l, we first note, given that we are assuming stationarity, that we can rewrite the joint distribution of the error terms in (3) in the form

$$\begin{aligned} f(\boldsymbol{\epsilon} | \sigma^2, \boldsymbol{\rho}) &= f(\epsilon_1) \textstyle\prod_{l=2}^{J} f(\epsilon_l | \epsilon_1, \ldots, \epsilon_{l-1}) \\ &= \textstyle\prod_{l=1}^{J} N(\nu_l,\ \sigma^2 K_l), \end{aligned} \tag{6}$$

where $\nu_l = E[\epsilon_l|\epsilon_1, \ldots, \epsilon_{l-1}]$, $l = 2, \ldots, m$ and K_l is a function of $\boldsymbol{\rho}$ (Harvey, 1993, pg 53). In a stationary AR(m) process $\nu_1 = 0$ and $(\nu_l,\ K_l) = (\sum_{k=1}^{m} \rho_k \epsilon_{l-k},\ 1)$ for $l = m+1, \ldots, J$. In order to find the remaining ν's and K's one can make use of the Durbin-Levinson algorithm (Brockwell and Davis, 1996, pg 67). In the case of an AR(2) process, for example,

$$\begin{gathered} K_1 = (1-\rho_2)/[(1+\rho_2)\{(1-\rho_2)^2 - \rho_1^2\}]^{-1}, \\ K_2 = (1-\rho_2^2)^{-1}, \\ \text{and} \\ \nu_2 = \rho_1/(1-\rho_2). \end{gathered} \tag{7}$$

For an AR(1) process simply set $\rho_2 = 0$ in (7).

It is immediately apparent, due to the fact that an AR(m) process is a Markov process, that each component of $\boldsymbol{\epsilon}$ is dependent only on its m nearest neighbors. Using this fact, the full conditional distribution of ϵ_l for the Gibbs sampler can be written as a function of the conditional normal distribution of ϵ_l given its m nearest neighbors. Therefore, the full conditional distribution for ϵ_l, $l = 1, \ldots, J$, is

$$f(\epsilon_l|\boldsymbol{\epsilon}_{-l}, \boldsymbol{\beta}, \sigma^2, \boldsymbol{\rho}, \mathbf{r}, D) \propto V \cdot \mathcal{L}_\phi \cdot \prod_{j=l}^{l+\min\{m,\ J-l\}} N(\nu_j,\ \sigma^2 K_j),$$

which, for each ϵ_l, can be condensed to the following form,

$$f(\epsilon_l|\boldsymbol{\epsilon}_{-l}, \boldsymbol{\beta}, \sigma^2, \boldsymbol{\rho}, \mathbf{r}, D) \propto V \cdot \mathcal{L}_\phi \cdot N\left(\mu_l/\eta_l,\ \sigma^2/\eta_l\right),$$

by completing the square. For an AR(2) error process

$$\mu_l = \begin{cases} \rho_1\epsilon_2 + \rho_2\epsilon_3 & l = 1 \\ \rho_1(\epsilon_1 + \epsilon_3) + \rho_2(\epsilon_4 - \rho_1\epsilon_3) & l = 2 \\ \rho_1(1-\rho_2)(\epsilon_{l-1} + \epsilon_{l+1}) + \rho_2(\epsilon_{l-2} + \epsilon_{l+2}) & l = 3, \ldots, J-2 \\ \rho_1(\epsilon_J + \epsilon_{J-2}) + \rho_2(\epsilon_{J-3} - \rho_1\epsilon_J) & l = J-1 \\ \rho_1\epsilon_{J-1} + \rho_2\epsilon_{J-2} & l = J \end{cases}$$

and

$$\eta_l = \begin{cases} 1 & l = 1 \text{ and } J \\ 1+\rho_1^2 & l = 2 \text{ and } J-1 \\ 1+\rho_1^2+\rho_2^2 & l = 3, \ldots, J-2 \end{cases}.$$

Once again, in order to obtain the full conditionals in the case of an AR(1) model, simply set $\rho_2 = 0$.

Due to the stationarity constraint on the autocorrelation parameters, $\boldsymbol{\rho}$, we must consider the joint full conditional distribution of $\boldsymbol{\rho}$ instead of assuming independent priors on the components of $\boldsymbol{\rho}$. Using the decomposition of $f(\boldsymbol{\epsilon}|\sigma^2, \boldsymbol{\rho})$ in (6), we can write the full conditional distribution of $\boldsymbol{\rho}$ as

$$f(\boldsymbol{\rho}|\boldsymbol{\beta}, \boldsymbol{\epsilon}, \sigma^2, \mathbf{r}, D) = f(\boldsymbol{\rho}|\boldsymbol{\epsilon}, \sigma^2) \propto \left(\prod_{j=1}^{m} K_j\right)^{-1} \exp\left\{-\frac{1}{2\sigma^2}\sum_{j=1}^{J}(\epsilon_j - \nu_j)^2/K_j\right\}\pi(\boldsymbol{\rho}).$$

The full conditional distribution of σ^2 is nearly identical to that of $\boldsymbol{\rho}$. Using the decomposition of $f(\boldsymbol{\epsilon}|\sigma^2, \boldsymbol{\rho})$, the full conditional of σ^2 is

$$f(\sigma^2|\boldsymbol{\beta}, \boldsymbol{\epsilon}, \boldsymbol{\rho}, \mathbf{r}, D) = f(\sigma^2|\boldsymbol{\epsilon}, \rho) \propto \sigma^{-J} \exp\left\{-\frac{1}{2\sigma^2}\sum_{j=1}^{J}(\epsilon_j - \nu_j)^2/K_j\right\}\pi(\sigma^2),$$

which is the form of an inverse gamma distribution with shape and scale parameters $J/2+1$ and $C(\boldsymbol{\epsilon}, \boldsymbol{\rho})/2 = \sum_{j=1}^{J}(\epsilon_j - \nu_j)^2/2K_j$, respectively. Therefore, if $\pi(\sigma^2)$ is an inverse gamma distribution with parameters a_0 and b_0, $\Gamma^{-1}(a_0,\ b_0)$, then the resulting conditional is an inverse gamma distribution with parameters $J/2 + a_0 + 1$ and $C(\boldsymbol{\epsilon}, \boldsymbol{\rho})/2 + b_0$. The full conditional of σ^2 is the only standard density.

When implementing Bayesian methodology, it is necessary to choose priors for the parameters. It is a standard practice in generalized linear models with random effects to assign the vague priors $\pi(\beta_l) = N(0, 1/\tau)$ for $l = 1, \ldots, P$ and $\pi(\sigma^2) = \Gamma^{-1}(\varepsilon, \varepsilon)$ where τ and ε are small (Dey et al., 2000, pg 400). In past Bayesian capture-recapture analyses, $\pi(r_l)$ has been chosen to be a beta distribution for $l = 1, \ldots, J$ (Brooks et al., 2000a) of which the uniform distribution is a special case for vague prior information. All of these priors can be easily modified to produce informative priors as desired.

When there is little or no prior information concerning the parameter $\boldsymbol{\rho}$, a uniform distribution on the region of stationarity would be the obvious choice for a noninformative prior distribution. This uniform distribution, however, may produce marginal priors which are not as vague as the researcher would like. In the AR(2) case for example, a uniform distribution for the AR parameters ρ_1 and ρ_2 over the region of stationarity produces marginal

distributions which are not uniform. In addition, a majority of the mass for the marginal distribution of ρ_2 will be located over negative values, producing a prior mean which is negative. This problem can often occur when building priors for parameter vectors over a constrained space. Barnard, McCulloch, and Meng (2000) illustrate the same dilemma when constructing priors for positive-definite covariance matrices.

In previous analyses using AR processes, the prior for the AR parameters was taken to be uniform over the stationary space of the parameters or a normal distribution if stationarity was not a concern or possibility (Huerta and West, 1999). Informative priors can also be constructed by truncating a multivariate normal to the stationary space. There is another approach, suggested by Sun and Berger (1998), that is useful for constrained parameter spaces. If we are concerned with the parameter vector $(\theta_1,\ \theta_2)$, then a prior can be built in the form $\pi(\theta_1)\pi(\theta_2|\theta_1)$. Using this method, we can often build a sufficiently noninformative prior that has better marginal properties. For example, in the AR(2) model, if we take $\pi(\rho_2) = U(-1, 1)$ and $\pi(\rho_1|\rho_2) = U(-(1-\rho_2),\ 1-\rho_2)$, we obtain a prior that approximates a joint uniform with marginal distributions centered on 0. The partial information approach can also be used to specify informative priors for some of the AR parameters, while leaving others vague.

Another practical aspect for the Bayesian analysis of AR(m) capture-recapture models is that a modified Gibbs sampler must be used due to the non-standard conditional distributions. In the following example, a Metropolis within Gibbs sampler (Gelman and Rubin, 1993) was used. Instead of successively sampling from the full conditional distributions to obtain a sample from the joint posterior, an observation is first drawn from a proposal distribution and then either accepted or rejected with a given probability.

4 Example: Northern Pintails

In order to illustrate the fitting of an AR(m) model to capture-recapture data we applied the Gibbs sampler methodology to a Northern Pintail band recovery data set for females in California. These data (Table 1) were first analyzed by Franklin et al. (2002) as part of a meta-analysis on long-term trends in avian survival for many North American bird species. The previous analysis was performed using a linear trend model, an identity link

function, and independent yearly random effects. The trend parameters as well as the variance component were estimated using the shrinkage estimation method of Burnham (2000). The slope estimate from the previous analysis is 0.0023 with an estimated standard error of 0.0051. The variance component is estimated to be 0.212.

The previous analyses detected no significant trend to survival probabilities over time. We will include a slope parameter in this example, however, as an illustration of the use of covariates in our estimation procedure. Therefore, we will use the model

$$\text{logit } \phi_j = \beta_0 + \beta_1(j - 14) + \epsilon_j \quad j = 1, \ldots, 27 \tag{8}$$

to illustrate the application of an AR(m) model. In this example, the covariate vector $\mathbf{X}_j'$ in (3) is given by $(1,\ j - 14)$. The time index is centered to reduce correlation of the β_1 sample with the β_0 sample, which leads to better exploration of the posterior density for each variable. In addition, since there seems to be no significant trend based on the previous analysis, we also analyzed the data without a slope parameter.

We chose to estimate separate reporting probabilities, λ_j's, for each year. Barry et al. (2001) note that separate λ_j's in (2) tend to confound the effects of a random survival process and this has been our experience as well. However, we have adopted a conservative strategy for making inference about a random survival process, by allowing for fluctuating reporting rates.

For this example, we have chosen the fit an AR(2) model to the data. This implies that the error terms in (8) follow the stochastic process

$$\epsilon_j = \rho_1\epsilon_{j-1} + \rho_2\epsilon_{j-2} + z_j, \qquad j = 1, \ldots, 27$$

The second order AR model was chosen based on a correlogram of the maximum likelihood estimates, using (2), of yearly survival probabilities from the program MARK (White and Burnham, 1999). By examining the correlogram we are treating the MLE survival estimates as time series data instead of estimates of time series data. So, if there is insignificant correlation at certain lags, it is likely that the corresponding AR coefficients will also be insignificant when they are simultaneously estimated with the covariate parameters.

The logit link function was chosen due to the fact that it is the most commonly used link in capture-recapture models. Capture-recapture data usually are not detailed enough

Table 1

Northern Pintail recovery data for banding years 1955 - 1983. The R_i represent the number of banded ducks released each year. Birds were banded in January of each Banding Year

Banding		Year of Recovery																										
Year	R_i	55	56	57	58	59	60	61	62	63	64	65	66	67	68	69	70	71	72	73	74	75	76	77	78	79	80	81	82
55	270	7	6	3	2	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0
56	693	21	10	4	2	3	0	0	0	0	0	1	0	0	1	0	0	0	0	0	0	0	0	0	0	0	0	0	0
57	1612	32	20	8	5	1	2	0	2	1	1	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0
58	858	26	12	5	6	4	1	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0
59	1471	21	18	6	5	0	0	1	0	1	1	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0
60	1051	18	4	6	4	1	2	0	1	1	0	0	1	0	0	1	0	0	0	0	0	0	0	0	0	0	0	0	0
61	796	24	6	4	0	3	3	0	0	0	1	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0
62	277	10	9	6	6	4	1	2	4	2	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0
63	903	15	8	1	8	4	0	2	1	1	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0
64	621	6	4	1	6	1	1	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0
65	584	10	4	3	7	3	1	2	1	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0
66	822	25	6	10	4	4	2	0	0	2	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0
67	1344	28	27	8	11	3	1	4	1	2	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0
68	566	10	13	6	2	2	1	1	1	1	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0
69	481	9	7	3	2	0	1	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0
70	695	11	11	5	2	2	1	1	0	1	1	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0
71	632	22	10	2	4	0	1	2	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0
72	1114	21	11	8	3	5	3	2	1	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0
73	639	9	10	10	2	3	0	2	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0
74	926	16	9	9	2	5	1	2	1	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0
75	858	14	12	3	5	1	1	1	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0
76	369	13	2	4	4	1	1	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0
77	450	8	3	4	1	2	1	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0
78	212	6	0	0	1	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0
79	1680	18	28	8	4	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0
80	421	14	1	2	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0
81	118	2	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0
82	60	1	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0

to detect subtle differences in the shape of the link used to constrain the survival probability to (0, 1). Even in the logistic regression scenario, it is often hard to distinguish between different link functions. For probabilities in the range 0.1 to 0.9, McCullagh and Nelder (1989, pg 109) note that it is difficult to discriminate between probit and logit links based on goodness-of-fit tests and for probabilities near 0.5, all four of the common links for binary data are close to one another. Survival probabilities are usually not near the extremes of 0 or 1 for North American duck species, so, it is reasonable to use the logit link function for these data.

The Bayesian software winBUGS (Spiegelhalter et al., 2000) was used to select the MCMC sample from the posterior distribution of $(\beta_0, \rho_1, \rho_2, \sigma, \boldsymbol{\epsilon})$. As was mentioned previously, there is only one parameter in which the full conditional is a standard density, therefore, a hybrid Gibbs sampler must be used. To accomplish this, winBUGS uses a Metropolis within Gibbs sampler where the proposal distribution is a normal distribution in which the variance adapts over the first 4,000 iterations to obtain an acceptance rate between 20% and 40%.

The priors chosen for the parameters were as follows:

$$(\beta_0, \beta_1)^T \sim N(\mathbf{0},\ 1/0.01\ \mathbf{I}), \qquad \sigma^{-2} \sim \Gamma(0.001,\ 0.001),$$
$$\rho_2 \sim U(-1,\ 1), \qquad \rho_1|\rho_2 \sim U(-(1-\rho_2),\ 1-\rho_2), \qquad \text{and}$$
$$\lambda_j \sim \text{i.i.d. } U(0,1) \quad j = 1, \ldots, 28.$$

These values were chosen to be sufficiently vague in order to induce little prior knowledge. The joint distribution of ρ_1 and ρ_2 was constructed to be a vague density over the region of stationarity with mean (0,0). A vague gamma distribution was chosen for σ^{-2} in order to take advantage of the standard full conditional distribution of σ^2.

In order to select the sample, two independent chains of 15,000 iterations each were run following a burn-in period of 5,000 iterations to allow the normal proposal distribution to finish adapting. The chains appeared to have converged well before the end of the burn-in period. Figure 1 shows gaussian kernel density estimates of the marginal posterior distributions for each of the parameters. These results suggest that, although the posterior density of ρ_1 seems to be centered directly over 0, the majority of the posterior mass for ρ_2 seems to be located over negative values indicating that there seems to be a significant influence on the error terms at the second lag. This suggests that if survival is high in one year, it will

Table 2

Posterior means, standard deviations, and 90% highest probability density (HPD) intervals for the AR(2) model parameters.

Model	Parameter	Mean	St. Dev.	90% HPD* Interval
Intercept and slope	β_0	0.600	0.159	(0.390, 0.850)
	β_1	-0.007	0.026	(-0.046, 0.033)
	ρ_1	0.014	0.288	(-0.458, 0.485)
	ρ_2	-0.452	0.307	(-0.928, 0.004)
	σ	0.688	0.222	(0.336, 1.015)
Intercept only	β_0	0.612	0.140	(0.409, 0.857)
	ρ_1	0.014	0.288	(-0.483, 0.456)
	ρ_2	-0.452	0.307	(-0.918, -0.109)
	σ	0.644	0.201	(0.330, 0.950)

* Estimated according to the algorithm presented by Chen et al. (2000).

be low in 2 years (lag 2). In addition, the posterior mass of σ seems to be located well away from 0, indicating that there is also a significant amount of random variation from year to year. The intercept parameter β_0 is also significantly greater than 0, which increases survival above approximately 0.5 on average. The posterior distribution of the slope parameter, β_1 appears to be highly concentrated near 0. While not directly comparable, the trend parameters and variance component are in qualitative agreement with the previous analysis. Figure 1 also illustrates the robustness of the marginal parameters to the presence or absence of the slope parameter. The marginal density estimates remain virtually unchanged. Posterior means, standard deviations, and 90% highest probability density (HPD) interval estimates are given in Table 2. The confidence intervals and approximate expected values support the conclusions that there exists a significant amount of variation not explained by the linear trend. There is also a high posterior probability that the slope parameter is approximately 0 and the second AR parameter is less than 0. In addition, since we have simulated values of $\boldsymbol{\epsilon}$ as well, we can estimate yearly survival as well. Figure 2 shows a plot of yearly survival with a 90% HPD confidence interval band for the intercept-only model.

The posterior densities remained virtually unchanged when the prior for the AR parameters is given by $\rho_1 \sim U(-2,\ 2)$ and $\rho_2|\rho_1 \sim U(-1,\ 1-|\rho_1|)$ as opposed to the priors used

previously. For these priors, the parameter estimates and corresponding 90% HPD intervals for the intercept-only model were β_0: 0.606 (0.395, 0.811), ρ_1: 0.018 (-0.505, 0.516), ρ_2: -0.544 (-0.915, -0.185), σ: 0.617 (0.299, 0.921). The posterior distributions of the AR parameters seem to be robust to different noninformative priors.

5 Discussion

Software to implement the methodology described here for an AR(2) band recovery model is available at no charge at `www.stat.colostate.edu/~jah/`. It is relatively straightforward to modify and implement this software for specific problems. The software is written in winBUGS, software for the Bayesian analysis of statistical models using Markov chain Monte Carlo methods, which is available at no charge at `www.mrc-bsu.cam.ac.uk/bugs`.

Bayesian methodology allows for time-series modeling of capture-recapture data not previously available. In addition to the univariate time-series models considered here, the random effects models could also be expanded to allow for other forms of dependence. For example, it might be of interest to model recapture or recovery rates with AR random effects. In that case, the recovery or recapture parameters are treated the same as the survival procedures presented. Another example is the consideration of gender in survival models. Common practice is to include gender as a covariate. In an AR model, this would imply that unknown environmental factors have the same effect on survival of males and females in each year and the level of association of survival across time remains the same between males and females as well. This may be an unrealistic assumption, so, it might be wise to model separate AR errors for males and females. To account for correlated errors between sexes, a multivariate AR process could be used with very little modification to the models proposed here.

Even though AR models can provide additional insight to the survival process, there are situations where estimation of the AR parameters may prove difficult. First, if there is fewer than 20 capture occasions in the data, there may not be enough data to greatly alter the prior distributions of the AR parameters. This is a problem often encountered in time series analysis. Secondly, if the recovery/recapture rates are very small there may be insufficient data to estimate AR parameters as well as covariate parameters. This second problem is

common to all capture-recapture data analysis. Finally, if there is very little error variation, an AR model is unlikely to provide any additional information. This last situation is really not a problem though, since a biologist's goal is usually to model survival with covariates. If all of the error in the survival process is accounted for, one can be confident of having a good description of the survival process. The AR models are implemented to account for unobserved environmental variation.

Some implications of using AR models with capture-recapture data is that the estimate of survival for any time period will have larger uncertainty than the simple covariate model. This variability is controlled by both variability of the white noise term in (4) and the AR parameters. For example, for an AR(1) model each random effect has a variance of $\sigma^2/(1-\rho^2)$. For σ held fixed, the variance of the random effect can get vary large as $|\rho| \to 1$. One can also observe, that for a fixed noise variance, the AR models will have larger variance than the independent random effects model.

One extension of the methodology described here is model selection. In general, model section is not an easy task for capture-recapture data in a Bayesian framework. Recently, King and Brooks (2001a; 2001b) have explored using Reverse Jump MCMC procedures (Green, 1995) for capture-recapture models with multiple strata and integrated recovery/recapture models. This provides the most promising solution, but, these procedures are not easily implemented for analysis on a regular basis. Another solution for Bayesian model selection is the Deviance Information Criterion (DIC) (Spiegelhalter et al., 2001). A nice feature of DIC is that the MCMC sample selected for parameter estimation can be used to construct a DIC score. Current formulations of the DIC, however, do not allow for distinguishing between different order AR processes for the capture-recapture models described here. If one wishes to fit an AR model to capture-recapture data, an initial step to select an appropriate order for the AR model is to fit a independent random effect model then plot a correlogram of the random effect point estimates and choose the order based on the plot. This will provide a conservative order for the model.

Overall, these models have the potential of providing wildlife biologists new insights into factors affecting survival for animals studied via capture-recapture studies.

Acknowledgments

The authors are grateful to Dr. David Anderson and Dr. Kenneth Burnham for suggestions and advice. The research was partially funded by NSF grant DUE-0094514 and EPA grant R-82909501-0.

References

Barnard, J., McCulloch, R., and Meng, X. L. (2000). Modeling covariance matrices in terms of standard deviations and correlations, with application to shrinkage. *Statistica Sinica*, 10:1281–1311.

Barry, S. C., Brooks, S. P., Catchpole, E. A., and Morgan, B. J. T. (2001). The analysis of ring recovery data using random effects. Preprint available at http://www.statslab.cam.ac.uk/∼steve/research.

Brockwell, P. J. and Davis, R. A. (1996). *Introduction to Time Series and Forecasting*. Springer, New York.

Brooks, S. P., Catchpole, E. A., and Morgan, B. J. T. (2000a). Bayesian animal survival estimation. *Statistical Science*, 15:357–376.

Brooks, S. P., Catchpole, E. A., and Morgan, B. J. T. (2002). "Bayesian methods for analysing ringing data" In Statistical Analysis of Data from Marked Bird Populations. *Journal of Applied Statistics*, 29(1-4):187–206. B.J.T. Morgan and D.L. Thomson, eds.

Brooks, S. P., Catchpole, E. A., Morgan, B. J. T., and Barry, S. C. (2000b). On the Bayesian analysis of ring recovery data. *Biometrics*, 56:951–956.

Brownie, C., Anderson, D. R., Burnham, K. P., and Robson, D. S. (1985). Statistical inference from band recovery data - a handbook. United States Department of the Interior Fish and Wildlife Service, Resource Publication 156.

Buckland, S. T., Goudie, I. B. J., and Borchers, D. L. (2000). Wildlife population assessment: past developments and future directions. *Biometrics*, 56:1–12.

Burnham, K. P. (2000). On random effects models for capture-recapture. Preprint available from author, Colorado State University.

Burnham, K. P. and White, G. C. (2002). "Evaluation of some random effects methodology applicable to ringing data" In Statistical Analysis of Data from Marked Bird Populations.

Journal of Applied Statistics, 29(1-4):245-262. B.J.T. Morgan and D.L. Thomson, eds.

Chen, M.-H., Shao, Q.-M., and Ibrahim, J. G. (2000). *Monte Carlo Methods in Bayesian Statistics*. Springer-Verlag, New York.

Cormack, R. M. (1964). Estimates of survival from the sighting of marked animals. *Biometrika*, 51:429–438.

Dey, D. K., Ghosh, S. K., and Mallick, B. K. (2000). *Generalized Linear Models: A Bayesian Perspective*. Marcel Dekker, Inc., New York.

Franklin, A. B., Anderson, D. R., and Burnham, K. P. (2002). Estimation of long-term trends and variation in avian survival probabilities using random effects models" In Statistical Analysis of Data from Marked Bird Populations. *Journal of Applied Statistics*, 29(1-4):267-289. B.J.T. Morgan and D.L. Thomson, eds.

Gelman, A. and Rubin, D. B. (1993). *in* Discussion on the meeting on the Gibbs sampler and other MCMC methods. *Journal of the Royal Statistical Society- Series B*, 55:73.

Green, P. J. (1995). Reversible jump Markov Chain Monte Carlo computations and Bayesian model selection. *Biometrika*, 82:711–732.

Harvey, A. C. (1993). *Time Series Models*. The MIT Press, second edition.

Huerta, G. and West, M. (1999). Priors and component structures in autoregressive time series models. *Journal of the Royal Statistical Society - B*, 61:881–899.

King, R. and Brooks, S. P. (2001a). Bayesian model discrimination for multiple strata capture-recapture. Preprint available at http://www.statslab.cam.ac.uk/∼steve/research.

King, R. and Brooks, S. P. (2001b). Models selection for integrated recovery/recapture data. Preprint available at http://www.statslab.cam.ac.uk/∼steve/research.

Lebreton, J., Burnham, K. P., Clobert, J., and Anderson, D. R. (1992). Modeling survival and testing biological hypotheses using marked animals: a unified approach with case studies. *Ecological Monographs*, 62:67–118.

Lindsey, J. K. (1999). *Models for Repeated Measurements*. Oxford University Press Inc., New York.

McCullagh, P. (1983). Quasi-likelihood functions. *The Annals of Statistics*, 11:59–67.

McCullagh, P. and Nelder, J. (1989). *Generalized Linear Models*, volume 37 of *Monographs on Statistics and Applied Probability*. Chapman & Hall/CRC, 2nd edition.

Poole, D. and Zeh, J. E. (2002). Estimation of adult bowhead whale survival rates using Markov chain Monte Carlo methods. *Journal of Agriculture, Biological, and Environmental Statistics*, 7:1–13.

Robert, C. P. and Casella, G. (1999). *Monte Carlo Statistical Methods*. Springer-Verlag, New York.

Spiegelhalter, D., Thomas, A., and Best, N. (2000). winBUGS version 1.3, User Manual. MRC Biostatistics Unit, Institute of Public Health, Cambridge UK.

Spiegelhalter, D. J., Best, N. G., Carlin, B. P., and van der Lind, A. (2001). Bayesian measures of complexity and fit. *Journal of the Royal Statistical Society -B*, to appear. Preprint available at http://www.mrc-bsu.cam.ac.uk/publications/preslip.shml.

Sun, D. and Berger, J. O. (1998). Reference priors with partial information. *Biometrika*, 85:55–71.

Vounatsou, F. and Smith, A. F. M. (1995). Bayesian analysis of ring recovery data via Markov Chain Monte Carlo simulation. *Biometrics*, 51:687–708.

White, G. C. and Burnham, K. P. (1999). Program MARK: survival estimation from populations of marked animals. *Bird Study*, 46 suppl:120–139.

Zeger, S. L. and Karim, M. R. (1991). Generalized linear models with random effects: a Gibbs sampling approach. *Journal of the American Statistical Association*, 86:79–86.

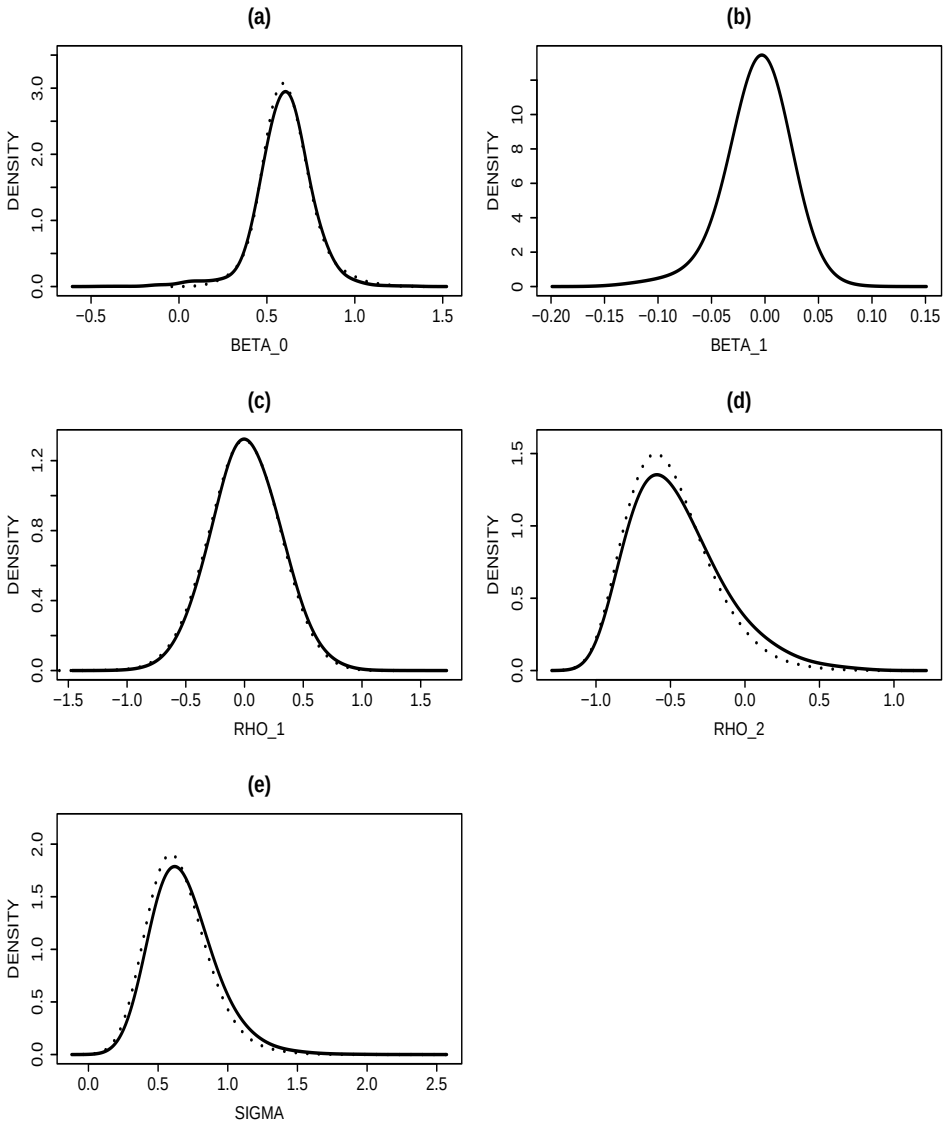

Figure 1. Marginal Posterior densities for (a): β_0, (b) β_1, (c): ρ_1, (d): ρ_2, and (e): σ from the Pintail data. The solid lines represent posterior densities from the time trend model, while the dotted lines represent the posterior densities when the trend parameter is absent.

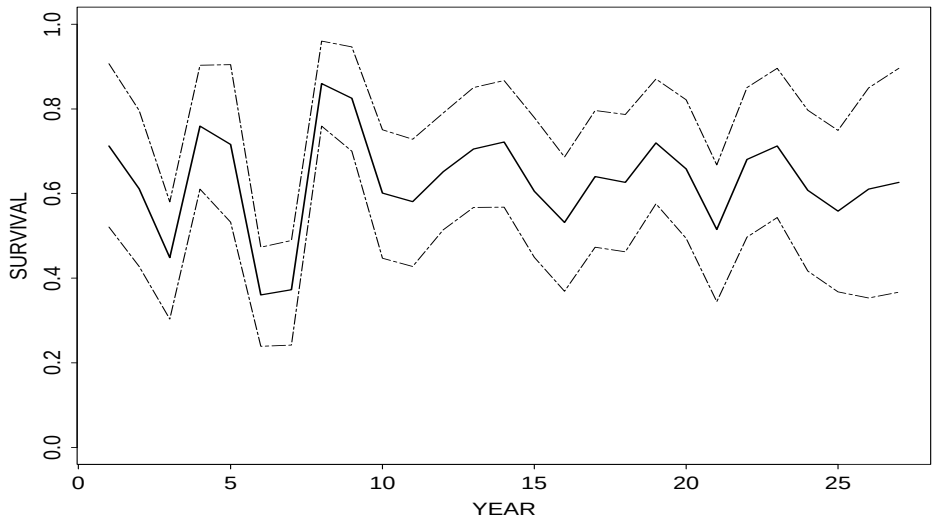

Figure 2. Plot of yearly survival estimates for Northern Pintail dataset with no linear time trend. The solid line is the estimated posterior mean survival and the dashed lines represent a 90% HPD interval.